Multimedia Systems
Lecture – 33, 34

By

Dr. Priyambada Subudhi

Assistant Professor

IIIT Sri City

Need for Video Compression

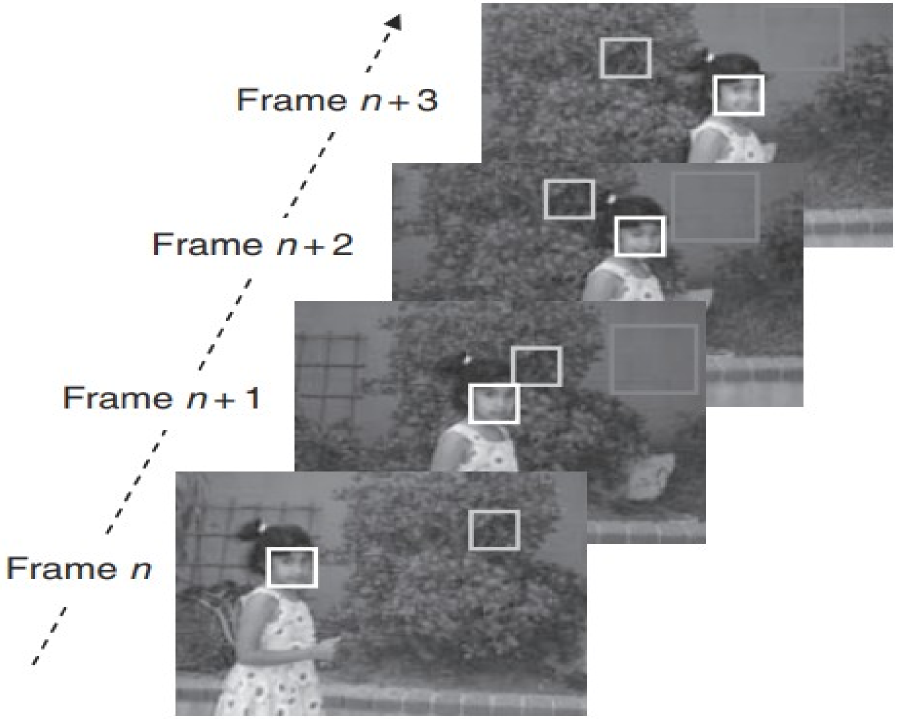
- HDTV broadcast has resolution of 1920×1080 at 30fps using 8bits to represent primary colors.
  - This leads to a total of 1.5 Gbps data rate
- But channel b.w. is only 6 MHz that supports around 19.2Mbps data rate only.
  - The channel has also to support audio and other data
  - Effective data rate reduced to only 18Mbps
- Therefore compression required is 1.5Gbps/ 18Mbps = 83
- Compression ratio required is 83:1

Basics of Video Compression

- Video consist of time-ordered sequence of frames, i.e. images.
- Images are having highly redundant data i.e. adjacent pixels are highly correlated.
- Image compression techniques mostly exploit both spatial and spectral redundancy.
- Image compression techniques like JPEG can be used to compress frames individually ?

- However, significantly higher compression rates can be achieved by exploiting another kind of redundancy in video—***temporal redundancy***.
- Just as pixel values are locally similar within a frame, they are also correlated across frames.
- For example, when you see a video, most pixels that constitute the backdrop or background do not vary much from frame to frame. Areas may move either when objects in the video move, or when the camera moves, but this movement is continuous and, hence, predictable.

Temporal Redundancy



Although each frame is different on a pixel-by-pixel basis, there is a great deal of correlation from frame to frame. The girl in the foreground moves, so the pixels on her face/dress have different locations in each frame, but the color values of these pixels do not change much.

- For video, it makes more sense to code only the changed information from frame to frame rather than coding the whole frame—akin to the ***predictive DPCM techniques.***
- The temporal correlation should result in the difference frame having lower information content than the frame itself.
- It would, therefore, be more efficient to encode the video by encoding the difference frames.
- Practically, however, the coherency from frame to frame is better exploited by observing that groups of contiguous pixels, rather than individual pixels, move together with the same motion vector. Therefore, it makes more sense to predict frame n +1 in the form of regions or blocks rather than individual pixels.

Frame difference

Frame n Frame $n+1$ Frame $(n+1) - $ Frame n

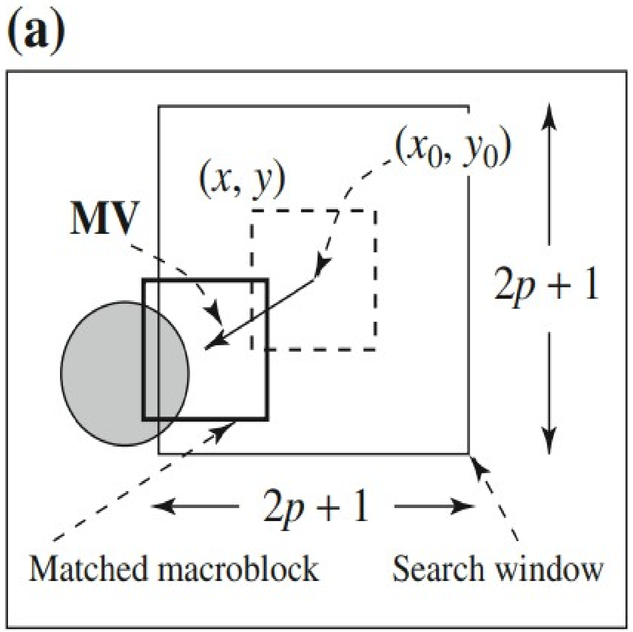
Key Idea

- Predict a new frame from the previous frame and specify the prediction error.
- Prediction error can be coded using image coding methods.
- Prediction from past frames is known as *forward prediction*.
- The current image frame, which needs to be compressed, is known as the *target frame*. The target frame is predicted on a block-by-block basis from a previous frame, which is known as the *reference frame.*
- Video compression makes use of *motion compensation* to predict a frame from the previous and/or next frame.

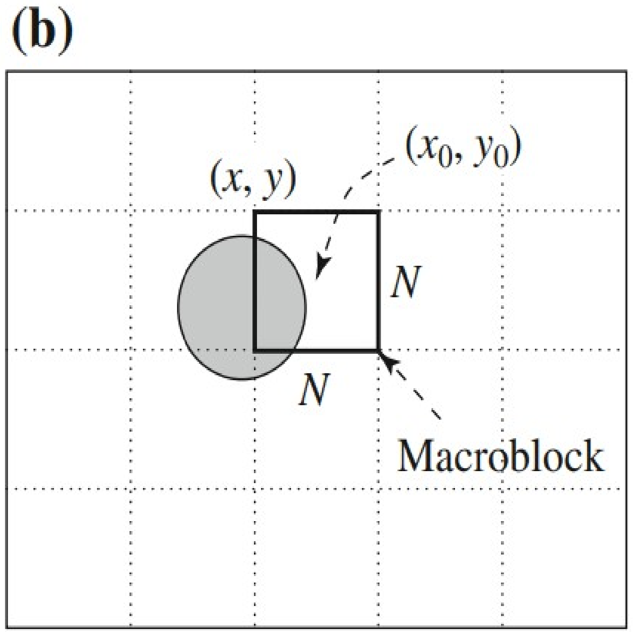
Key Terms

- ***Macro Blocks:*** Compression method works on a block of 16×16 pixels called macro blocks.
- ***I (intra) frames***
  - Independent frames
  - Coded without reference to other frames
  - I frames are interspersed periodically in the encoding process, and provide access points to the decoder to decode the succeeding predicted frames.
- ***Inter or P (Predictive) frame***
  - Not independent
  - These frames are predicted from a past frame (I or P)
  - Coded by forward predictive coding
    - Current macro block is predicted from similar macro block in the previous I or P frame and the difference between the macro blocks is coded.

Macroblocks and motion vector in video compression

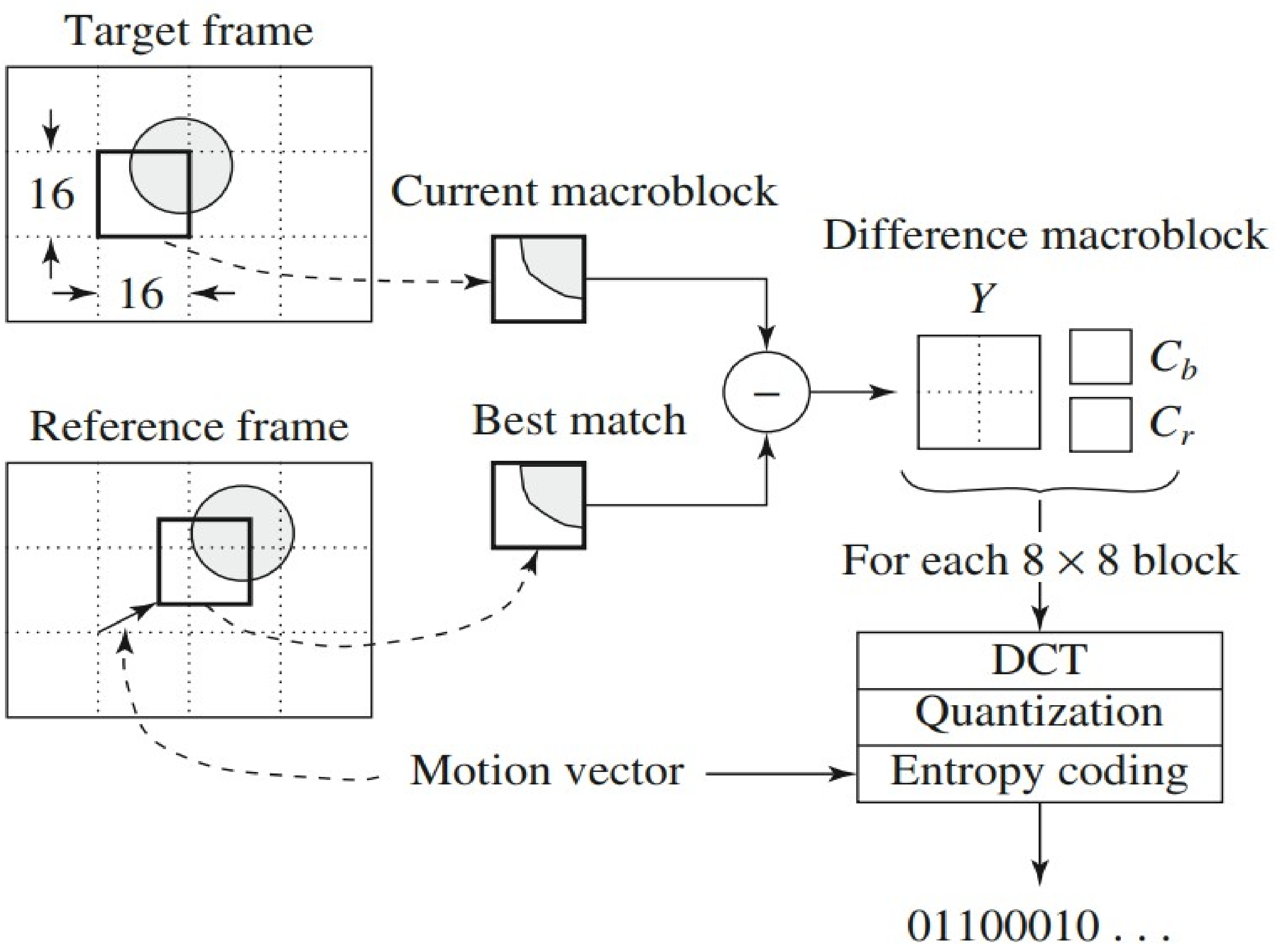


Reference Frame



Target Frame

P-frame coding based on motion compensation

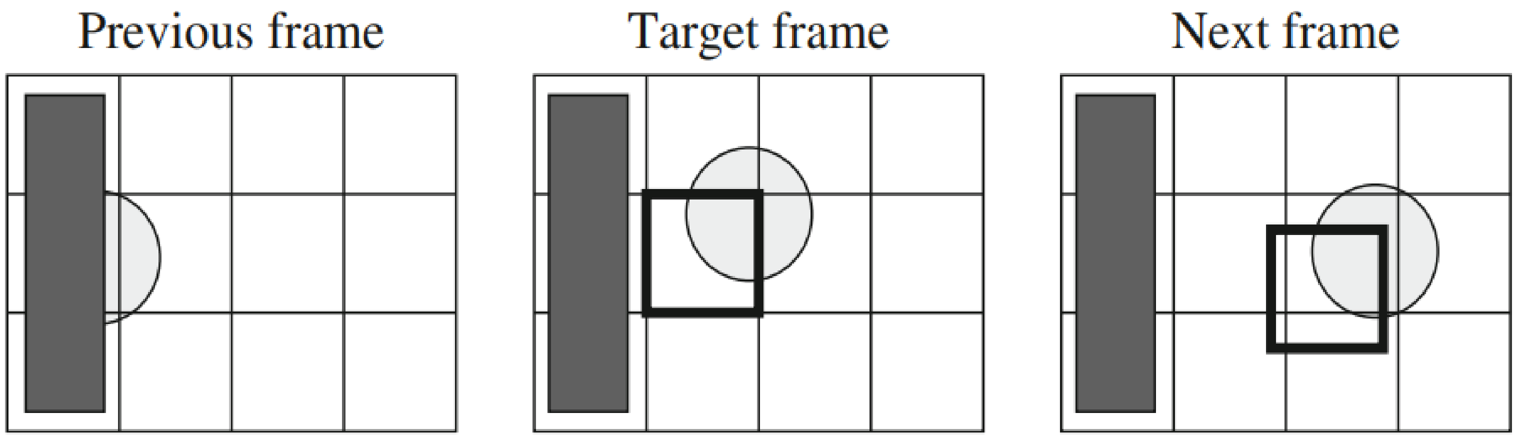


MPEG-1 Evolution

- Approved by the ISO/IEC MPEG group in November 1991 for Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s.
- Common digital storage media include ***compact discs (CDs)*** and ***video compact discs (VCDs).***
- MPEG-1 supports only ***noninterlaced video***.
- It uses ***4:2:0*** chroma subsampling.
- Audio-coding standard consists of three layers of audio-coding schemes with increasing complexity, resulting in successively better compression. These are better known as MPEG-1 Layer I, MPEG-1 Layer II, and MPEG-1 Layer III, popularly known as ***MP3***.

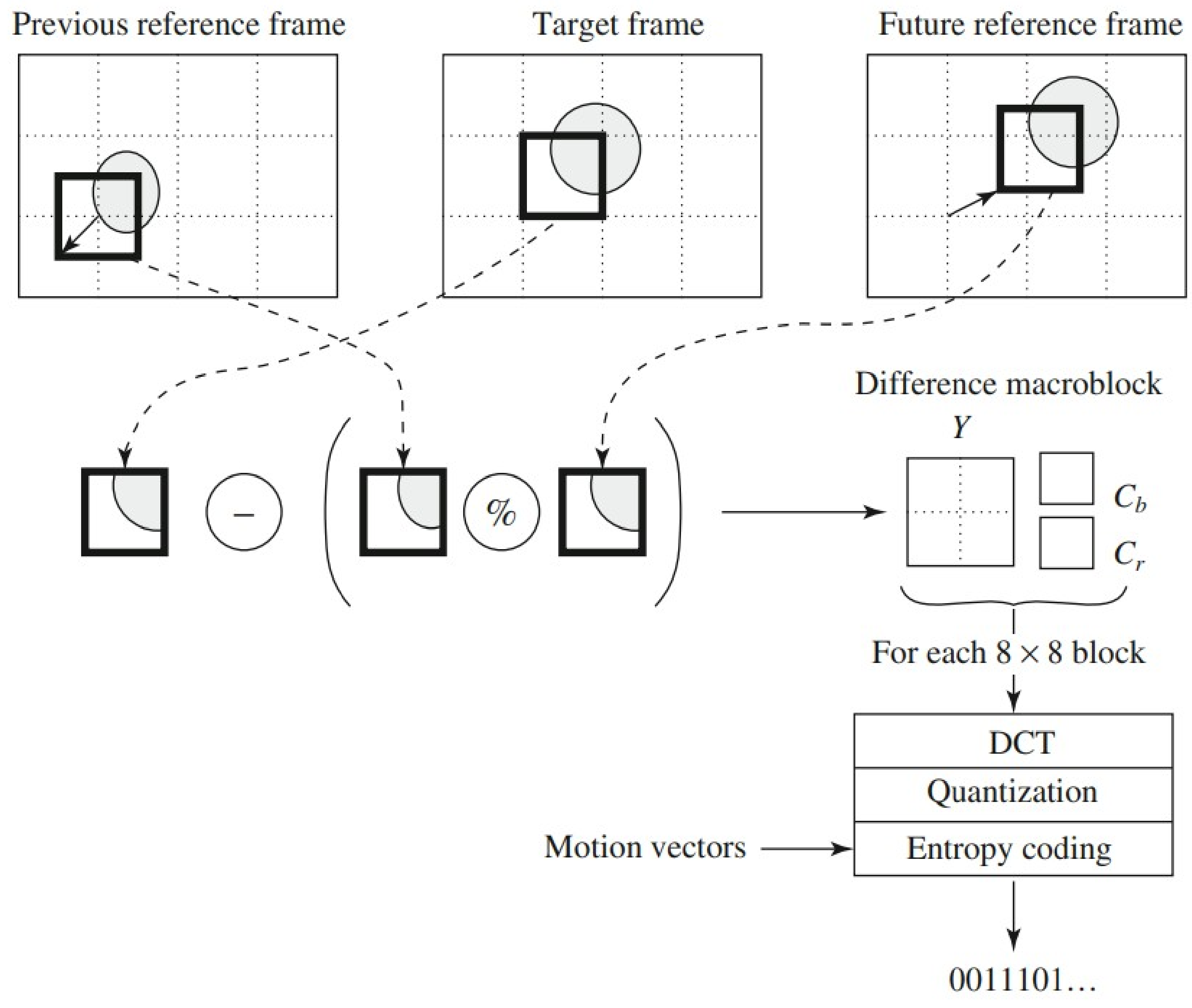
- ***MPEG*** introduces a new B-frame.
- ***B-frame (Bidirectional predictive coded)***
  - Due to unexpected movement in all real scenes, the target macro block may not have a good matching in the previous frame.
  - Therefore, B-frame is coded with reference to both previous and future reference frames (either I or P).
  - When prediction is from a previous frame, it is called forward prediction.
  - When the matching macroblock is obtained from a future I- or P-frame in the video sequence, then it is known as backward prediction.

The need for bidirectional search



The macroblock containing part of a ball in the target frame cannot find a good matching macroblock in the previous frame, because half of the ball was occluded by another object. However, a match can readily be obtained from the next frame.

Motion Compensation in MPEG-1

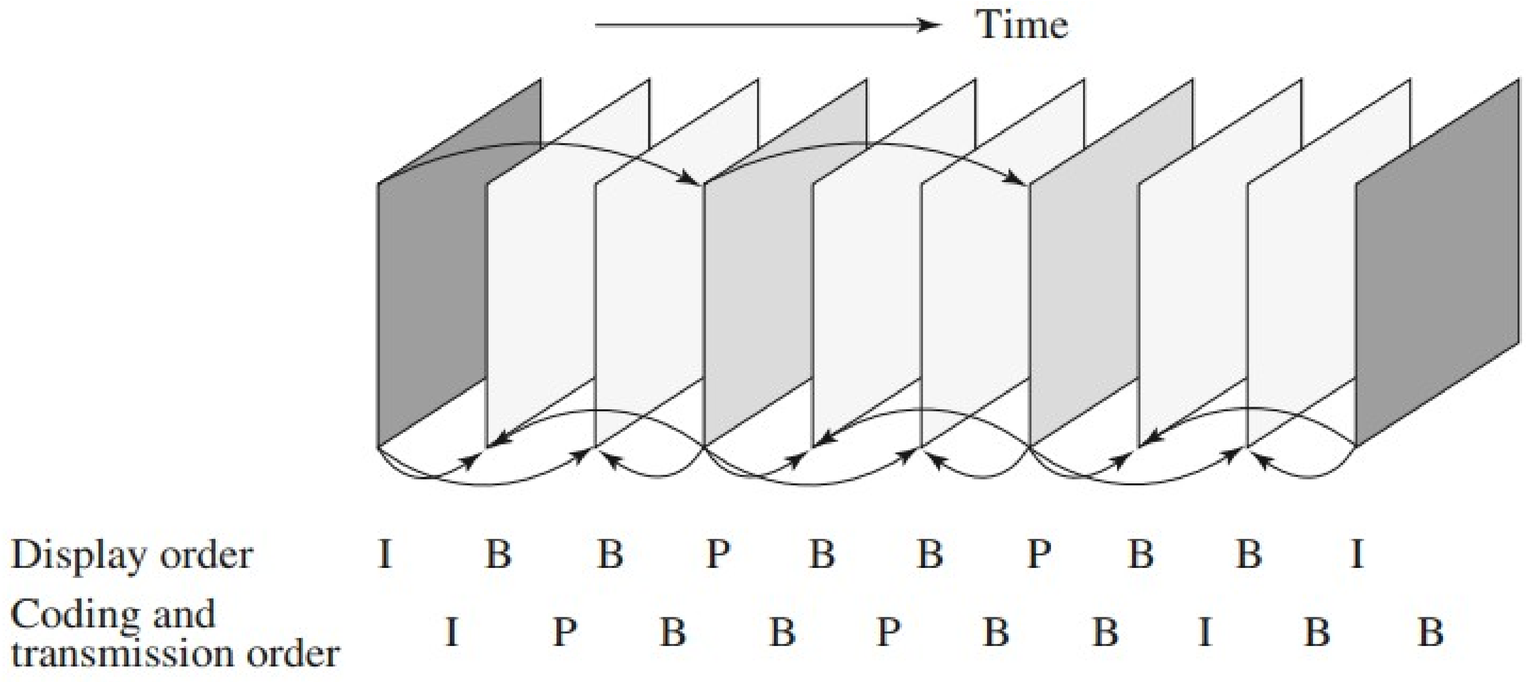


- If matching in both directions is successful, two motion vectors will be sent, and the two corresponding matching macroblocks are averaged (indicated by "%" in the figure) before comparing to the target macroblock for generating the prediction error.

- If an acceptable match can be found in only one of the reference frames, only one motion vector and its corresponding macroblock will be used from either the forward or backward prediction.

- B frames also necessitate reordering frames during transmission, which causes delays. The order in which frames arrive at the encoder is known as the ***display order***.

MPEG Frame sequence



The decoder has to have both the past and the future reference frames ready before it can decode the current B frame. Consequently, the encoder has to encode and send to the decoder both the future and past reference frames before coding and transmitting the current B frame. This enforces the encoder to make the coding and transmission order different from the display order.

- All potential B frames need to be ***buffered*** while the encoder codes the future reference frame, imposing the encoder to deal with buffering and also causing a ***delay*** during transmission.
- The actual frame pattern is determined at encoding time and is specified in the video's header. MPEG uses M to indicate the interval between a P-frame and its preceding I- or P-frame, and N to indicate the interval between two consecutive I-frames. Here, M = 3, N = 9.
- A special case is M = 1, when no B-frame is used.
- Issues with MPEG-1
- The inevitable delay and need for buffering become an important issue in real-time network transmission, especially in streaming MPEG video.
- Stores and plays video on the CD of a single computer at a low bitrate (1.5 Mbps)

MPEG-2

- Project started in 1990 and finalized in 1994.
- Provided high quality video with bit rate around 4Mbps.
- It was developed as a standard for digital tv broadcast.
- It gained wide popularity in terrestrial, satellite, or cable network.
- Adopted in DVDs

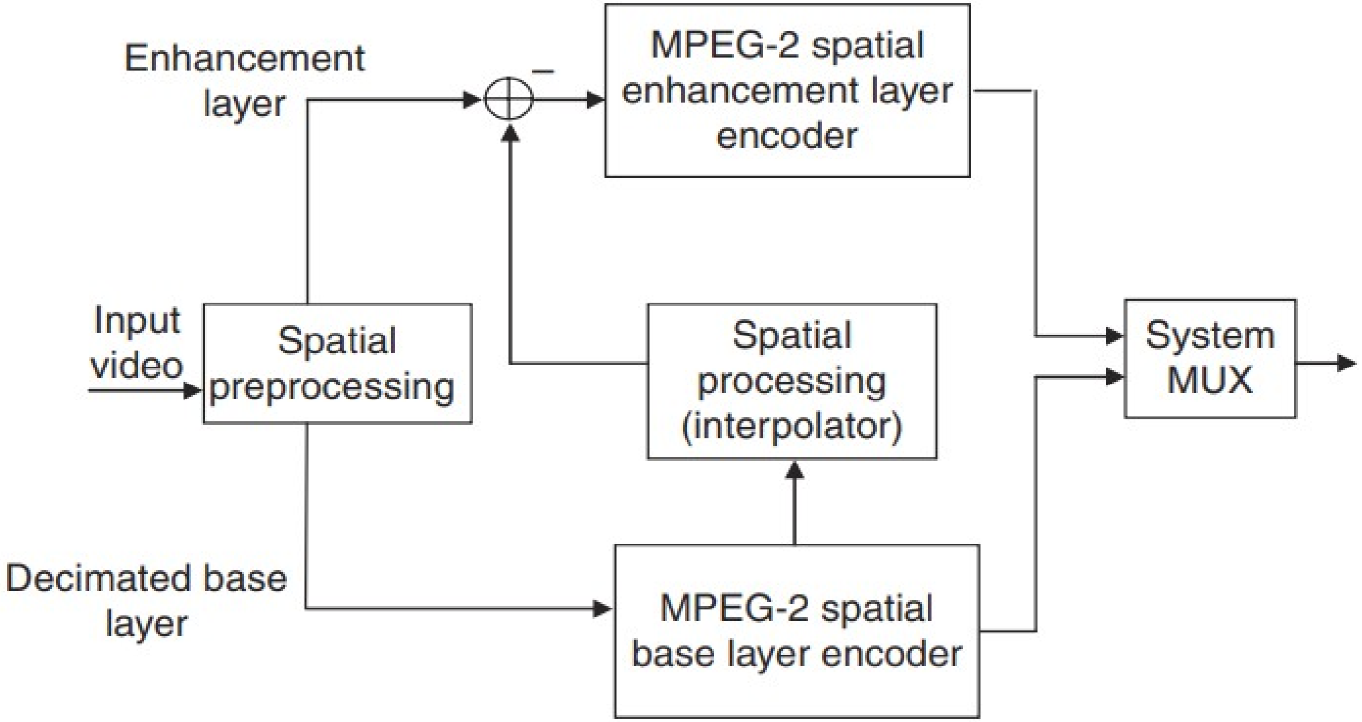
Salient features of MPEG-2 video coding

- MPEG-2 encodes video using I, P, and B frames similar to MPEG-1 with ***halfpixel approximation*** for motion vectors.
- MPEG-2 has support for interlaced video that consists of alternating fields.
  - During intermode compression, fields need to be predicted from fields. For this, MPEG-2 defines different modes of prediction for predicting frames from frames, fields from fields, and fields from frames.
- In addition to defining video and audio, MPEG-2 was also designed as a transmission standard providing support for a variety of ***packet formats*** with error-correction capability across noisy channels.

- One new aspect introduced in MPEG-2 video was ***scalable encoding***, which generates coded bit streams in such a way that decoders of varying complexities can decode different resolutions and, thus, different quality from the same bit stream.

- This allows for encoders to encode once and support different bandwidths, having end clients with decoders of different complexity.

Scalable MPEG-2 video encoding



The input video is broken into two layers, a base layer containing low frequency and an enhancement layer containing high frequency. Both layers are coded independently and then multiplexed into the coded bit stream.

MPEG-4

- MPEG-4 is one of the latest encoding standards from the MPEG community and has a much broader scope than just encoding video or audio.
- Finalized and ratified in 1999; however, the later versions such as MPEG-4 version 10 were completed later in 2003.
- Version 10 of the visual part is also known as ***MPEG-4 AVC*** (Advanced Visual Coding)
- MPEG-4 arose from a need to have a scalable standard that supports a wide bandwidth range from low bandwidth to high bandwidth.

Salient features of MPEG-4

- It supports both ***progressive and interlaced*** video encoding.
- The video compression obtained by MPEG-4 ***ASP (Advanced Simple Profile)*** is better than MPEG-2 by ***25%*** for the same video quality. This is because MPEG-4 does ***quarter-pixel accuracy*** during motion prediction.
- The standard is ***object based*** and supports multiple video streams that can be organized (along with other audio, graphics, image streams).
- MPEG-4 supports coding video into ***video object planes***. A video can consist of different video object planes (***VOP***s), and the image frames in each VOP can have ***arbitrary shapes***, not necessarily rectangular. This is useful for sending two video feeds to the client, who can then composite them for a MPEG-4 player.

- For example, one video could be of a reporter in a studio chroma-keyed in front of a blue screen, while the other could be that of an outdoor setting. The composition of both occurs at the client side. This increases the flexibility in creating and distributing content.

- MPEG-4 compression provides ***temporal scalability***, which utilizes advances in object recognition. By separating the video into different layers, the foreground object, such as an actor or speaker, can be encoded with lower compression while background objects, such as trees and scenery, can be encoded with higher compression.

Object-based coding in MPEG-4